Original Article

Optimal sperm length for high siring success depends on forehead patch size in collared flycatchers

Murielle Ålund, Siri Persson Schmiterlöw, S. Eryn McFarlane, and Anna Qvarnström
Department of Ecology and Genetics, Animal Ecology, Uppsala University, Norbyvägen 18D 75236 Uppsala, Sweden

Received 6 December 2017; revised 17 June 2018; editorial decision 15 July 2018; accepted 23 July 2018; Advance Access publication 9 August 2018.

INTRODUCTION

It is now well recognized that sexual selection does not stop at mating. Instead, in most taxa, it continues in the form of sperm competition between ejaculates of rival males (Parker 1970), due to widespread female polyandry, and mechanisms of cryptic female choice (Eberhard 1996). A complete understanding of sexual selection thus requires a combination of traditional studies of sexual display traits and intra- and intersexual behaviors, detailed analyses of traits involved in postcopulatory sexual selection and contributing to fertilization success, as well as investigating the relationships and potential trade-offs between episodes of pre- and postcopulatory selection (Simmons and Fitzpatrick 2016; Simmons et al. 2017). Despite decades of studies on this subject in various taxa and across very different mating systems, the link between pre- and postmating sexual selection is still unclear (Griffith et al. 2002; Kvarnemo and Simmons 2013; Travers et al. 2016).

Should we expect male display traits to reveal information about the quality of their sperm? A male’s reproductive success depends on his ability to secure mating opportunities with a fertile female, to maintain high fertilization success and minimize sperm competition with other males (e.g. through mate guarding behaviours and/ or “aggressive” ejaculates) and to some extent through successful extra-pair copulations or sneaking behaviours (Whittingham and Dunn 2004; Blengini et al. 2014; Fitzpatrick et al. 2016). In case of energy limitation, we can easily imagine that investment in pre- and postmating traits trade-off against each other (Parker 1998; Evans 2010). Alternatively, successful males might be better at all tasks and advertise high fertilization efficiency through their secondary sexual characters (the phenotype-linked fertility hypothesis, Sheldon 1994).

Estimating sperm “quality” is difficult, because ejaculate characteristics conferring high fertility differ between species, depending on the fertilization environment (internal or external) and on population-specific conditions such as the level of sperm competition (Johnson et al. 2013; Fitzpatrick and Lupold 2014; Simpson et al. 2014). Both high sperm numbers and high swimming velocity are important contributors to fertilization success, but the relative importance of sperm numbers versus specific successful...
sperm characteristics differs between taxa (Immler et al. 2011). Sperm form and size are extremely diverse and often highly divergent between species, suggesting an important function in fertilization success (Pitnick et al. 2009). The relationships between sperm length, swimming velocity, and fertilization success are complex and vary between taxa, and sometimes even between populations (Fitzpatrick and Lupold 2014), but longer sperm are generally found to be faster, particularly in birds (Lüpold et al. 2009). In populations of captive zebra finches, sperm size and swimming speed are genetically codetermined (Mossman et al. 2009) and sperm morphology is repeatable and heritable (Birkhead et al. 2005; Kim et al. 2017; Knief et al. 2017) in the zebra finch. Recent competitive experiments in this species have also shown long sperm to have higher fertilization success (Bennison et al. 2015). More generally, sperm swimming speed and longevity or storage duration have been found to trade-off against each other in birds, insects and mammals (e.g. Immler et al. 2007; Helfenstein et al. 2010; Ramon et al. 2013), suggesting that shorter sperm might be more beneficial for long-term survival in female storage organs. High levels of sperm competition have recently been shown to accelerate divergence in sperm traits between passerine species (Rowe et al. 2015). Sperm competition is also expected to reduce variation in sperm phenotypes within species in various taxa, by eliminating abnormal sperm and selecting towards an optimal phenotype (Hunter and Birkhead 2002; Calhim et al. 2007; Firman and Simmons 2010; Fitzpatrick and Baer 2011). High variation both between males within species and within single ejaculates are however regularly reported (Fitzpatrick et al. 2010; Ramón et al. 2014; Rowe et al. 2015), and the maintenance of high variation in sperm morphology despite sperm competition is puzzling.

Thus, despite numerous efforts in trying to understand the relationship between primary and secondary sexual characters, results are inconsistent between studies (Mautz et al. 2013; Lüpold et al. 2015; Mautz and Ålund 2016) and links between display traits and sperm traits, if found, are very weak (Förstmeier et al. 2017). Determining which traits influence mating and fertilization success is often complicated by a lack of direct measures of siring success or any knowledge of the male’s condition and breeding status. Though in vitro fertilization trials or controlled mating experiments in captivity are very useful to determine the properties and functions of different types of sperm, actual scenarios of sperm competition in the wild are more complex, including interactions with multiple social partners or rivals, limited energy availability, and intricate suites of behavioural strategies for all involved individuals.

In this study, we focus on a wild population of collared flycatchers, a passerine bird that has been subject to numerous studies on premating sexual selection. Males compete over breeding territories, using the size of their white forehead patch as a badge of status, indicating dominance (Part and Qvarnström 1997). This secondary sexual character is also used by females for mate choice and known to influence the probability of obtaining extra-pair copulations (Sheldon et al. 1997; Qvarnström et al. 2006), as well as the level of paternal care provided by the male (Qvarnström et al. 2000; Qvarnström 2003). Here, we investigate whether and how the size of this sexual signal and sperm morphology interact in determining siring success among naturally paired birds over 4 breeding seasons. We report context-dependent effects of sperm size on fertilization efficiency, where the optimal sperm size, in terms of siring success, depends on the size of the males’ secondary sexual character. Males with relatively small forehead patches sire more offspring when they have relatively long sperm, but the opposite pattern is found among males with large forehead patches. This implies that males use different mating and fertilization strategies depending on their levels of experience, dominance status and sexual attractiveness, likely in response to very different sperm competition pressures.

**METHODS**

**Study species and long-term monitoring**

Collared flycatchers (*Ficedula albicollis*) are small, migratory passerine birds that overwinter in sub-Saharan Africa and breed in Europe between April and July (Qvarnström et al. 2010). They started colonizing Oland, in the Baltic Sea (Sweden) in the 1960’s, most likely after a founder event from the neighbouring island, Gotland (Kardos et al. 2017). They breed in deciduous forests and feed primarily on flying insects and caterpillar larvae (Vallin et al. 2011; Qvarnström et al. 2016; Rybinski et al. 2016). Females are brown-gray and males are black and white, with a white collar on their neck and a white forehead patch. Our study population has been monitored since 2002, across several forest areas located in the central and northern parts of Oland, using over 2000 nest-boxes. All breeding adults and nestlings are individually marked, measured and sampled for blood every year. The male white forehead patches are measured to the nearest mm using callipers. We measure the width and height of the male forehead patches and compute the forehead patch area as the product of these 2 measurements, which is used in all models below and referred to as “forehead patch size” hereafter. Laying date, clutch size, fledgling success and recruitment rates are also recorded for all breeding pairs (Qvarnström et al. 2010). All procedures on collared flycatchers were approved by the Linköping Animal Care Board (Jordbruksverket—Linköpings Djurförsökssetiska DNR 21-11).

**Sperm samples and analyses**

We sampled sperm from 131 males caught between 10 May and 24 June for 4 consecutive years (2010–2013), including 14 individuals caught multiple times, using traps while they defended a nestbox (9 males) or were feeding nestlings (41 males) or using playback and mist nets for nonterritorial males (60 males), or while their social females were laying eggs or incubating (21 males). We define 2 categories of males in our dataset: “breeding males” that were attending a nest (usually feeding nestlings) at some point during the season, and “non-breeding males,” that never were found building or attending a nest, but were only caught earlier in the season when trying to attract females to their territories. Both breeding and non-breeding males were sampled throughout the breeding season (median sampling day: breeders: 8 June, nonbreeders: 9 June). Nearly all breeding males sampled started breeding during the “peak of laying,” loosely defined as the 6–10 days period during which 80% of the birds in the population laid their first egg that same year (usually in mid-May; calculated for a total of 589 breeding records from 2010 to 2013). This is the period where sperm competition is expected to be highest, when most females are laying eggs and thus potentially available for extra-pair copulations.

Ejaculates were collected through cloacal protuberance massaging (Wolfson 1952) and stored in 5% formalin. Samples were scanned for intact sperm cells drop by drop (9 μL at a time). In 2010, we used an Olympus BX11 microscope, and took 5 pictures per male at ×400 magnification, using a Nikon digital sight DS-2Mv camera (resolution 2Mpix) and the Nis-Elements imaging software for Nikon (FPackage, 1991–2009). For samples from 2011 to 2013,
we photographed and measured 10 sperm per male using a Nikon ECLIPSE Ci microscope (Nikon DS-Qi1Mc) mounted with a Nikon DIGITAL SIGHT DS-U3 camera. Different parts of the sperm (head, midpiece, and tail) were measured using the software ImageJ [ImageJ 1.41, Wayne Rasband, National Institute for Health, USA, rsb.info.nih.gov/ij/) in 2010 and NIS-Elements Basic Research 3.22.11 in 2011–2013. The sperm total length was computed as the sum of all components. Identity of the person measuring sperm morphology and type of microscopic and software used were directly confounded with year (controlled for in the models) and did not affect the results when tested separately (results not shown).

Paternity analyses

We performed paternity analyses on 425 nestlings from 71 different nests with known mothers and social fathers (caught feeding the nestlings). The average clutch size in our population was 6, ranging from 4 to 8 eggs per clutch. We used the software Cervus (v. 3.0.3, copyright Tristan Marshall 1998–2007, www.fieldgenetics.com, Kalinowski et al. 2007) to compare nestlings and their putative fathers at 10 to 13 microsatellite loci (FhU1, FhU2, FhU3, FhU4, PdoU5, Fhy304, Fhy401, Fhy403 and Fhy454, with the addition of Fhy407 in 2010–2012, and Fhu223, Fhy223, and Fhy330 in 2011–2013, Haavie et al. 2000; Leder et al. 2008). Following well-described protocols (e.g. Jones et al. 2010; Cramer et al. 2016), data was simulated for 10000 offspring with 5 candidate fathers, assuming that 70% of the population had been sampled. We only included individuals that could be compared using at least 6 microsatellite loci. We considered nestlings to be sired by an extra-pair male when the confidence level for the pairwise comparison between the offspring and the social father was <95%.

Statistical analyses

In order to test the reliability of the measurements, repeatabilities (Lessels and Boag 1987; Nakagawa and Schielzeth 2010) were calculated using 5 sperm from 6 males measured 4 times each by each measurer. The repeated measures were performed haphazardly, allowing various periods of time between measurements of the same cell and blind to previous measurements. The repeatability estimates of the different sperm parts for 2011–2013 is given here, with estimates for 2010 (different measurer) in brackets: sperm head: 0.948 (0.78), midpiece: 0.995 (0.97), tail: 0.989 (1.0).

We tested for a relationship between male phenotype, breeding status and sperm length using a linear mixed effects model, with the total sperm length as a response variable (repeated measures of 5–10 sperm per male) and forehead patch size, age (fitted as a quadratic variable), and breeding status (2-level factor, breeding or not) with all possible interactions as fixed effects. We fitted male individual identity (to account for repeated measures of several sperm per individual), year, and the identity of the person measuring the bird’s forehead patch (i.e. “forehead patch measurer”) as random effects. Sperm length and forehead patch size were both fitted as z-standardized variables (value minus the population’s mean, divided by the standard deviation) to ease comparison between these variables, as advised by Schielzeth (2010) and Grueber et al. (2011). We chose sperm total length as the response variable in order to capture the whole variation in sperm size among and within ejaculates (see below for details about other sperm components).

We estimated sperm performance using a generalized linear mixed effects model with siring success as a response variable, computed as a binary vector (number of offspring sired versus not sired in a male's nest), using the clind function in lme4 (see below) and setting the family to “binomial” (logit). We fitted sperm length, forehead patch size (both z-standardized, see above), age (quadratic variable), the interactions between these 3 variables and clutch size as fixed effects, and year and patch size measurer as random effects. This subset of our dataset, with information on siring success, did not contain any repeated males. We note that principal component analyses did not help capture the variation in sperm design in our dataset (results not shown) and report separate models (to avoid overfitting our models) using each sperm component as a predictor of siring success in the supplementary material (Supplementary Tables S1–S3), as well as false-discovery-rate adjusted P values to account for multiple testing (Forsman and Schielzeth 2011).

All statistical analyses were done in R (version 3.3.2) (R Core Team 2016) using RStudio version 1.0.136, (RStudio Team 2016). Between-year repeatability in sperm length was calculated using the package “rptR” (Stoffel et al. 2017) with the lmz method and 1000 bootstrap samples. We used the packages “lme4” (Bates et al. 2014) and “lmerTest” (Kuznetsova et al. 2014) that implement Satterthwaite’s approximations of degrees of freedom (Satterthwaite 1946) for linear mixed effects models.

RESULTS

We measured sperm morphology using ejaculates sampled on 131 occasions from 117 individual males caught during 4 consecutive breeding seasons (2010–2013), including 104 adults (2–9 years old) and 27 yearlings. Of all the birds measured, 71 were caught feeding nestlings. One of them did not sire any offspring in his own nest and thus constituted an outlier in our dataset. This individual was thus removed from analyses on siring success. We measured 5 sperm per male in 2010 and 10 sperm per male in 2011–2013. The average lengths of the different sperm components over all sampled individu-

als were as follows (range with standard deviation in brackets): total length: 101.38 µm (90.4–109.59, SD = 2.93), head: 11.07 µm (9.65–15.18, SD = 1.03), midpiece: 74.75 µm (67.06–81.48, SD = 2.43), tail: 15.87 µm (7.23–23.93, SD = 2.97). The variance in sperm total length between males was 5.73 (CI [2.3, 8.5]), while the within-male variance was 3.00 (CI [1.15–5.94]), i.e. 35% of the total variance. 14 birds were caught in more than one year. This allowed the calculation of between-year repeatability of sperm length (Nakagawa and Schielzeth 2010; Stoffel et al. 2017), which was 0.65 (CI [0.329, 0.853]).

Male phenotype and sperm morphology

We used a linear mixed effects model to test for covariance between the birds’ phenotype (age, forehead patch, breeding status, and all possible interactions as fixed effects) and the length of their sperm, using repeated measures of 5–10 sperm per male. We included an individual’s identity, year, and the forehead patch measurer’s identity as random effects. We find a significant 3-way interaction between forehead patch size, age, and breeding status on sperm length (F1221 = 4.29, P = 0.04, Table 1, Figure 1). Middle-aged males have longer sperm than both juvenile and old males, except for breeding males with relatively small forehead patches. In this category, middle-aged males by contrast have the shortest sperm. The average sperm lengths and forehead patch sizes did not differ between breeding and nonbreeding males (t-tests, P = 0.93 and 0.11, respectively, forehead patch sizes (mm): breeding males: mean = 82.23, range = 51.66–141.52, SD = 15.83; nonbreeding males: mean = 85.85, range = 52.56–115.56, SD = 13.56, see above for details on sperm length).
We performed paternity analyses on 421 nestlings from 70 different nests with known mothers and social fathers (the subset of our sampled males, previously referred to as "breeding males"). In total, 17.1% (72) of the nestlings were sired by an extra-pair male, in 45.7% (32) of the surveyed nests. There was a significant difference in sperm length between males that sired all of the nestlings in their nests and males that lost some paternity to other males ($t$-test, $t_{63.99} = 2.22, P = 0.029$). Males that sired all of the offspring in their nests had on average shorter sperm. We used a generalized linear model to test the relationship between a male’s average sperm length and forehead patch size (z-standardized predictor variables) on his siring success (i.e. the proportion of offspring he sired). The bird’s individual identity, year, and the forehead patch measurer’s identity were included as random effects. Age was fitted as a quadratic variable in this model. We found a 3-way interaction between forehead patch size, breeding status, and age. When males are split into forehead patch size categories for illustration purposes (see Figure 1) middle-aged males have the longest sperm, except for breeding males with relatively small forehead patches. In this category, middle-aged males by contrast have the shortest sperm.

Figure 1

The relationship between male age and sperm length differs depending on the males’ pairing status and the size of their display trait (i.e. forehead patch) in collared flycatchers. Middle-aged males have the longest sperm, except for breeding males with relatively small forehead patches. In this category, middle-aged males by contrast have the shortest sperm.

### Table 1

Results of a linear mixed effects model testing for relationships between male traits known to influence mating behaviors, that is, age, a sexual display trait (forehead patch size) and breeding status, and sperm length (repeated measures for 5–10 sperm per male) in collared flycatchers.

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Standard error</th>
<th>df</th>
<th>$t$ value</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.267</td>
<td>0.237</td>
<td>7.55</td>
<td>-1.124</td>
<td>0.295</td>
</tr>
<tr>
<td>Age (quadratic)</td>
<td>-0.008</td>
<td>0.044</td>
<td>141.02</td>
<td>-0.179</td>
<td>0.858</td>
</tr>
<tr>
<td>Standardized patch size</td>
<td>0.186</td>
<td>0.112</td>
<td>261.84</td>
<td>1.662</td>
<td>0.098</td>
</tr>
<tr>
<td>Breeding status</td>
<td>-0.058</td>
<td>0.143</td>
<td>233.94</td>
<td>-0.402</td>
<td>0.688</td>
</tr>
<tr>
<td>Age × standardized patch size</td>
<td>-0.009</td>
<td>0.039</td>
<td>186.85</td>
<td>-0.234</td>
<td>0.815</td>
</tr>
<tr>
<td>Age × breeding status</td>
<td>0.023</td>
<td>0.059</td>
<td>360.56</td>
<td>0.388</td>
<td>0.608</td>
</tr>
<tr>
<td>Standardized patch size × breeding status</td>
<td>-0.042</td>
<td>0.149</td>
<td>223.82</td>
<td>-0.285</td>
<td>0.776</td>
</tr>
<tr>
<td>Age × standardized patch size × breeding status</td>
<td>-0.207</td>
<td>0.100</td>
<td>132.21</td>
<td>-2.071</td>
<td>0.040</td>
</tr>
<tr>
<td>Random effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual</td>
<td>0.768</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measurer</td>
<td>0.128</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>0.023</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>0.649</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The bird’s individual identity, year, and the forehead patch measurer’s identity were included as random effects. Age was fitted as a quadratic variable in this model. We found a 3-way interaction between forehead patch size, breeding status, and age. When males are split into forehead patch size categories for illustration purposes (see Figure 1) middle-aged males have the longest sperm, except for breeding males with relatively small forehead patches. In this category, middle-aged males by contrast have the shortest sperm.

Forehead patch size, sperm morphology, and siring success

We performed paternity analyses on 421 nestlings from 70 different nests with known mothers and social fathers (the subset of our sampled males, previously referred to as “breeding males”). In total, 17.1% (72) of the nestlings were sired by an extra-pair male, in 45.7% (32) of the surveyed nests. There was a significant difference in sperm length between males that sired all of the nestlings in their nests and males that lost some paternity to other males ($t$-test, $t_{63.99} = 2.22, P = 0.029$). Males that sired all of the offspring in their nests had on average shorter sperm. We used a generalized linear model to test the relationship between a male’s average sperm length and forehead patch size (z-standardized predictor variables) on his siring success (i.e. the proportion of offspring he sired).
sired in his own nest, fitted as a binomial response variable). There was a significant interaction between a male’s sperm length and his forehead patch size on his siring success (2-way interaction, est = −0.402 ± 0.18, z = −2.18, P = 0.029, P_{adj} = 0.116 [Table 2, Figure 2a]). Males with relatively small forehead patch sizes sired more offspring within their social nests when they had longer sperm while the opposite was true for males with larger forehead patch sizes. We also report a marginal 3-way interaction between forehead patch size, sperm length, and age (as a quadratic variable) on siring success (P = 0.055, P_{adj} = 0.138), that improves the fit of the model ([Table 2, Figure 2b]). This model has a marginal R^2 of 0.11 (i.e. only looking at the fixed effects) and a conditional R^2 of 0.39 (i.e. fixed and random effects together explain about 39% of the variance).

**DISCUSSION**

In our monitored population of collared flycatchers, forehead patch size and sperm length interact in determining male siring success. Having relatively longer sperm, which often means faster swimming speed, particularly in birds (Lüpold et al. 2009; Bennison et al. 2015), is only beneficial for males with small forehead patches. In contrast, males with large forehead patches instead have a higher siring success when their sperm are relatively short (Figure 2a). Below, we discuss...

### Table 2

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>Estimate</th>
<th>Standard error</th>
<th>z value</th>
<th>P value</th>
<th>Adjusted P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>−1.042</td>
<td>1.273</td>
<td>−0.819</td>
<td>0.413</td>
<td>0.288</td>
</tr>
<tr>
<td>Standardized patch size</td>
<td>−0.223</td>
<td>0.190</td>
<td>−1.176</td>
<td>0.240</td>
<td>0.288</td>
</tr>
<tr>
<td>Standardized sperm length</td>
<td>−0.415</td>
<td>0.201</td>
<td>−2.056</td>
<td>0.040</td>
<td>0.08</td>
</tr>
<tr>
<td>Age (quadratic)</td>
<td>0.046</td>
<td>0.158</td>
<td>0.293</td>
<td>0.770</td>
<td>0.341</td>
</tr>
<tr>
<td>Clutch size</td>
<td>0.356</td>
<td>0.184</td>
<td>1.933</td>
<td>0.053</td>
<td>0.212</td>
</tr>
<tr>
<td>Standardized patch size × standardized sperm length</td>
<td>−0.402</td>
<td>0.184</td>
<td>−2.185</td>
<td>0.029</td>
<td>0.08</td>
</tr>
<tr>
<td>Standardized patch size × quadratic age</td>
<td>0.128</td>
<td>0.135</td>
<td>0.953</td>
<td>0.341</td>
<td>0.45</td>
</tr>
<tr>
<td>Standardized sperm length × quadratic age</td>
<td>−0.336</td>
<td>0.216</td>
<td>−1.558</td>
<td>0.119</td>
<td>0.159</td>
</tr>
<tr>
<td>Standardized patch size × standardized sperm length × quadratic age</td>
<td>0.574</td>
<td>0.299</td>
<td>1.919</td>
<td>0.055</td>
<td>0.138</td>
</tr>
</tbody>
</table>

**Random effects**

<table>
<thead>
<tr>
<th></th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurer</td>
<td>1.287</td>
</tr>
<tr>
<td>Year</td>
<td>0.221</td>
</tr>
</tbody>
</table>

We computed siring success as a binomial response variable including the number of sired versus nonsired offspring in each male’s nest. We fitted z-standardized forehead patch size, z-standardized sperm length (calculated from the average sperm length for each male), and the interaction between them as fixed effects, as well as clutch size. The forehead patch measurer’s and the year of sampling were included as random effects. We found a significant interaction between forehead patch size and sperm length on siring success, where males with relatively small forehead patches sire more offspring if they have longer sperm, whereas the opposite pattern is found among males with larger forehead patches. The 3-way interaction between forehead patch size, sperm length, and age (as a quadratic variable) on siring success, is reported here because it improves the fit of the model.

---

**Figure 2**

(a) Male collared flycatchers with relatively small display traits (i.e. forehead patches) sire more offspring if they have longer sperm, while the opposite pattern is found among males with larger forehead patches. Male collared flycatchers with relatively large display traits sire more offspring if they have shorter sperm. Males are split into categories of patch sizes (small-average-big) for representation purposes only and forehead patch size was used as a continuous variable in all analyses. (b) Middle-aged collared flycatcher males tend to have higher siring success, but this pattern seems to be reversed in males with average forehead patches and short sperm.
our findings in relation to current ideas about predicted relationships between display traits and sperm characteristics and argue that the concepts of optimal “sperm quality” need to be revised.

**General patterns of sperm morphology in collared flycatchers**

We report moderate variation in sperm length within and between collared flycatcher males, calculated for 117 different males, in a population with a proportion of extra-pair young (17.1% of nestlings) close to the average for passerine birds (Låfjeld et al. 2010). We found high between-year repeatability of sperm length, consistent with previous reports of repeatability and heritability of this trait in passerine birds (Birkhead et al. 2005; Mossman et al. 2009; Cramer et al. 2013). When investigating the relationship between male phenotype and sperm morphology in our population, we find that breeding status, age and the size of the white forehead patch all interact in predicting an individual’s sperm size. Specifically, middle-aged males have longer sperm than both juvenile and old males, except for breeding males with relatively small forehead patches. In this category, middle-aged males by contrast have the shortest sperm (Figure 1). This shows that sperm morphology varies both with male phenotypes important for securing territories and mating success and with the resulting breeding status, likely influencing behavioural strategies (including extra-pair copulations) and the risk of sperm competition.

**Forehead patch size, sperm morphology and siring success**

We find evidence that the benefits of a given sperm length, in terms of siring success, depend on the size of the secondary sexual character among 70 breeding males. Having relatively longer sperm is beneficial for males with small forehead patches in terms of higher relative fertilization success in their own nest, while the opposite is true for males with larger forehead patch size (Figure 2a). As mentioned above, male collared flycatchers with bigger forehead patches are more aggressive and dominant, and thus likely better at mate guarding, but also more likely to secure extra-pair copulations (Part and Qvarnstrom 1997; Sheldon et al. 1997). As a result, small-patched males are more likely to get cuckolded and face strong sperm competition. To maintain within-pair paternity, these males might have to invest more into “defensive” sperm characteristics, which make their sperm successful in direct competitive situations with sperm from other males. Differences in sperm characteristics of dominant and subdominant males have previously been reported in birds: the sperm of subdominant chicken swim faster (Froman et al. 2002), and subdominant males maintain high ejaculate quality over more mating events than dominant males do (Cornwallis and Birkhead 2007). Recent studies also found differences in ejaculate quality of house sparrows depending on their social status (Rojas Mora et al. 2017a, 2017b). Differences in sperm morphology have also been found between mating strategies, as sperm of superb fairy wrens with higher within-nest siring success had longer flagella (Calhim et al. 2011). Several fish species exhibit more extreme behavioural differences with males adopting different reproductive tactics depending on their phenotypes, and differences between the sperm of “guarder” and “sneaker” males have also been reported, as well as rapid changes in ejaculate properties following changes in sperm competition risk (Vladič and Jarvi 2001; Evans 2010; Egeland et al. 2015; Fitzpatrick et al. 2016; Bartlett et al. 2017). Here we show that even in a system with only moderate levels of sperm competition (17.9% of extra-pair nestlings) and less extreme hierarchical dominance, optimal “sperm quality” differs depending on male phenotype.

**Revisiting the idea of optimal “sperm quality”**

Previous studies have suggested that there should be stabilizing selection on sperm length in response to sperm competition, towards an optimal sperm phenotype in the population (Låfjeld et al. 2010; Laskemoen et al. 2013; Rowe et al. 2015). Our results, indicating that patterns of selection on sperm characteristics differ between males depending on their secondary sexual characters and potentially also on their age, could instead explain the lack of consistency among the numerous studies attempting to link primary and secondary sexual characters (Mautz et al. 2013; Lupold et al. 2015; Mautz and Ålund 2016). Our results imply that the most successful sperm phenotypes differ between males within the same population. Males with big ornaments achieve higher fertilization success when producing shorter sperm. This means that we can refute both a trade-off between sizes of primary and secondary sexual characters and the phenotype-linked fertility hypothesis. Indeed, if males with big ornaments seem to have enough energy to also produce longer sperm, this is not beneficial to them in terms of siring success. As longer sperm may swim faster and shorter sperm might live longer (Mossman et al. 2009; Helfenstein et al. 2010; Fitzpatrick and Lupold 2014), it is tempting to speculate about the exact mechanisms behind these results. Higher risk of cuckoldry might favour fast-swimming sperm for quick fertilization, while more attractive males could rely on long-lived (or more numerous, see Immler et al. 2011) shorter sperm to fertilize a whole clutch. Substantial within male variation means that males could also benefit from producing different types of sperm in order to increase their chances of fertilization in different contexts (see Helfenstein et al. 2010). By calculating actual fertilization success in the nest of each male, we identify the most successful sperm phenotypes for each category of males, after sperm competition and cryptic female choice, regardless of the relationships between sperm size and function. Studies looking for correlations between sexual signals and sperm quality often lack actual data on fertilization success and base their assumptions about sperm quality on results from other studies or systems, which might not always result in correct assessment of sperm traits. Additionally, a lack of knowledge on a male’s age and breeding status might result in the impossibility to detect relationships between investment in sexual signals and ejaculates or potential trade-offs in producing different types of ejaculates.

This study highlights the complexity of interactions within and between sexes in the wild, which cannot be captured solely by in vitro experiments testing for sperm competition, without any further knowledge about the males sampled. Similarly, even detailed studies of within and extra-pair copulation decisions and frequencies by both sexes do not translate in a full understanding of outcomes of sperm competition in terms of siring success (Girndt et al. 2018). Although specific sperm traits might confer a competitive advantage in controlled experiments, levels of sperm competition might be very asymmetric for different categories of males and different sperm traits may be optimal for fertilization depending on whether the female obtained the sperm from a preferred male or not. Even if males may gain reproductive success by siring extra-pair offspring, securing high within-pair paternity should be more important, particularly in socially monogamous species that depend highly on bi-parental care (Lebigre et al. 2012). Furthermore, successful extra-pair copulations might often be initiated by the
females and leave little control to the males (Girndt et al. 2018). Very dominant males might almost never face sperm competition and choose to invest more in behavioural tactics preventing their females from engaging in extra-pair copulations. Finally, the traits we investigated only explain a small proportion of the variation in siring success between males. Variation in seminal fluid proteins, whose functions are often still unknown, are likely to also affect fertilization success (Locatello et al. 2013; Yamane et al. 2015; Lewis and Pitcher 2017). Additionally and most importantly, interactions between male and female reproductive proteins and mechanisms of cryptic female choice (Cramer et al. 2016) are likely to strongly affect the outcome of sperm competition, even when overall successful sperm phenotypes have been identified (Chow et al. 2013; Bennison et al. 2015; Reinhart et al. 2015; Hemmings et al. 2016). Future studies integrating knowledge about female behaviours and reproductive biology will help understanding the complex interactions determining fertilization outcomes in the wild.

SUPPLEMENTARY MATERIAL
Supplementary data are available at Behavioral Ecology online.

FUNDING
This work was supported by the Swedish Research Council (VR 621-2012-3722 to A.Q.), the Stiftelsen for Zoologisk Forskning (M.A. and S.E.M.), the Swedish Ornithological Foundation (M.A.) and the Natural Sciences and Engineering Research Council (PGS-D to S.E.M.).

We thank Amber Rice, Päivi M. Sirkia, Emily R. A. Cramer, William Jones, and numerous field assistants for help with data collection and Reija Dufva for performing the molecular analyses.

Conflict of interest: We have no competing interests.

Data accessibility: Analyses reported in this article can be reproduced using the data provided by (Ålund et al. 2018).

Handling editor: John Fitzpatrick

REFERENCES


